



TECHNISCHE
UNIVERSITÄT
DARMSTADT

ULB

Registerbewusstsein schärfen, domänenadaptiv schreiben lernen

Kermes, Hannah

(2020)

DOI (TUprints): <https://doi.org/10.25534/tuprints-00017199>

Lizenz:



CC-BY-NC-ND 4.0 International - Creative Commons, Namensnennung, nicht kommerziell, keine Bearbeitung

Publikationstyp: Buchkapitel

Fachbereich: 02 Fachbereich Gesellschafts- und Geschichtswissenschaften

Quelle des Originals: <https://tuprints.ulb.tu-darmstadt.de/17199>

Registerbewusstsein schärfen, domänenadaptiv schreiben lernen

HANNAH KERMES

Die Twitter Nachricht im Internet, die Email an Freunde, der Zeitungstext, der literarische Aufsatz oder das Essay in der Schule: Die Vielzahl an verschiedenen Registern und Textsorten stellt heutzutage eine immer größere Herausforderung dar, sowohl in der Produktion als auch in der Rezeption. Um Nachrichten im Fernsehen, der Zeitung oder auf Twitter, (wissenschaftliche) Artikel und Fachbeiträge verstehen zu können, müssen wir wissen, wie Inhalt dort sprachlich kodiert ist. Auch für die Produktion müssen wir die jeweiligen Mechanismen, Terminologien und Sprachfloskeln beherrschen. Das Bewusstsein für Sprachregister muss aber geschärft und die benötigten sprachlichen Fähigkeiten trainiert werden. Dies gilt sowohl für die Muttersprache als auch für Fremdsprachen.

Digitale Medien können helfen, Sprachvariation sichtbar zu machen. Explorative Verfahren ermöglichen einen intuitiven Zugriff auf die in einem Text oder Register verwendeten sprachlichen Mittel, insbesondere auf die Abstraktion von der Mikrostruktur des einzelnen sprachlichen Signals auf die Makrostruktur des typischen Sprachmusters für den vorliegenden Text, das vorliegende Register.

Die Registerforschung beschäftigt sich schon lange mit der Untersuchung der sprachlichen Unterschiede und Gemeinsamkeiten von Register und Produktionstypen (Biber u.a. 1999; Biber 1995; Biber u.a. 2001; Quirk u.a. 1985; Zwicky u.a. 1982). Dabei kommen unterschiedliche Werkzeuge und Methoden aus der Korpuslinguistik, der Statistik aber auch aus dem Data Mining zum Einsatz, um relevante Merkmale zu identifizieren und zu untersuchen. Die Frage ist, inwieweit diese Methoden und Werkzeuge auch in der Fachdidaktik eingesetzt werden können, um das Bewusstsein und den richtigen sprachlichen Umgang mit Registern zu schärfen und zu erlernen.

Für den Einsatz in der Fachdidaktik bieten sich zunächst klassische Methoden und Werkzeuge aus der Korpuslinguistik, aber auch einfache Techniken aus dem Data Mining an. Frequenzdistributionen von lexikalischen Einheiten, Wortsequenzen und Wortarten, Kollokationen, Keywordanalyse, lexikalische Dichte und Varianz, Wortwolken typischer Wörter (basierend auf Sprachmodellen/ Häufigkeiten und Distanzmaßen) können eingesetzt werden, um von einzelnen sprachlichen Realisierungen auf Sprachmuster zu schließen, die für ein bestimmtes Register oder einen bestimmten Produktionstyp typisch ist. Sie bieten einen anderen Blick auf das sprachliche Signal, eine andere Abstraktionsebene, die Regelmäßigkeiten und Konventionen sichtbar und begreifbar macht.

Im Folgenden wollen wir zunächst kurz auf die Registerforschung eingehen und Methoden und Werkzeuge aus der Registeranalyse vorstellen, die auch für die Fachdidaktik geeignet sind. Anschließend wollen wir zeigen, wie diese Methoden in der universitären Lehre zum Einsatz kommen und Möglichkeiten für einen Einsatz in der Fachdidaktik diskutieren. Folgende konkrete Szenarien wollen wir dabei vorstellen:

- mit Wortwolken typischer Wörter bzw. Keywordanalyse sprachliche Unterschiede erkennen (Funktionswörter vs. Inhaltswörter, Funktionswörter als grammatische Merkmale)
- einfache Frequenzdistributionen auf Wortebene erstellen

Ein besonderes Augenmerk wird dabei auf dem fachdidaktischen Konzept und der Umsetzbarkeit liegen. Es werden einfach anzuwendende, möglichst webbasierte Werkzeuge vorgestellt, für die Fachdidaktik interessante Aspekte sowie die konkrete Vorbereitung bzw. Umsetzung eines solchen Konzeptes aufgezeigt. Dabei greifen wir auf Erfahrungswerte bei der Vermittlung solcher Konzepte u.a. für Lehramtsstudierende zurück.

5.1 Registerforschung

Unter Registervariation versteht man die situationsbezogene Variation einer Sprache. Die verschiedenen Situationen werden durch Diskursparameter gekennzeichnet: *field*, *tenor* und *mode*. *Field* bezieht sich dabei auf die inhaltliche Ausrichtung des Textes, *tenor* auf die Einstellung, den Standpunkt des Produzenten und Rezipienten und *mode* auf das Medium und die strukturelle Darstellung des Inhalts (vgl. Halliday 1985; Quirk u.a. 1985; Halliday u.a. 1985). Den Parametern sind unterschiedliche linguistische Subsysteme zugeordnet: *field* der Lexis und Colligation, *tenor* dem Modus und der Modalität und *mode* den strukturgebenden sprachlichen Mitteln wie Thema-Rhema oder gegeben-neu.

In einem ersten Schritt werden für diese Parameter anhand der Subsysteme konkrete sprachliche Merkmale identifiziert. Diese werden dann für die zu untersuchenden Register aus Korpora extrahiert und anschließend analysiert und interpretiert.

Die wohl bekannteste und am weitesten verbreitete Methode zur Untersuchung von sprachlicher Variation ist die *multidimensionale Methode (MD)* von Biber (1988; 1992). Die multidimensionale Methode untersucht Registervariation anhand großer Textcorpora mit Hilfe von computerlinguistischen Werkzeugen und multivariater statistischer Methoden (etwa der Faktoranalyse oder des Clustering). Multivariate Methoden sind nötig, weil die MD davon ausgeht, dass sprachliche Varianz in verschiedenen linguistischen „Dimensionen“ (Ebenen) stattfindet, die jeweils durch eine Gruppe sprachlicher Merkmale gekennzeichnet ist (z.B. Nomen und Präpositionalphrasen, Verben im *past* oder *present tense*¹). Abbildung 1 zeigt eine Zusammenfassung der von Biber (1992) aufgestellten Zuordnung sprachlicher Merkmale zu den fünf Hauptdimensionen sprachlicher Varianz hier für das Englische. Die Zahlenwerte in der Abbildung zeigen den Einfluss des jeweiligen sprachlichen Merkmals auf die sprachliche Dimension von -1.0 (uncharakteristisch) bis 1.0 (charakteristisch). Je weiter der Wert von 0 entfernt ist, desto aussagekräftiger ist das Merkmal.

Bei der MD werden die sprachlichen Merkmale für die spätere Analyse aus Korpora extrahiert und anschließend mittels Häufigkeitsverteilungen und multivariater statistischer Methoden analysiert. Die Extraktion erfolgt auf zuvor linguistisch annotierten Korpora. Die Annotation kann dabei unterschiedlich komplex sein. In der Regel sind die Korpora zumindest lemmatisiert und wortartengetaggt. Die Korpora können aber auch semantische, syntaktische oder diskursbezogene Annotationen enthalten. Auch die Wortartentags selbst können mehr oder weniger Information enthalten, etwa zu Tempus und Numerus, Aspekt, lexikalischen Klassen, etc.

Die extrahierten Daten werden dann für jeden Text bzw. für jedes Register in einem multidimensionalen Merkmalssatz gespeichert. Dieser multivariate Datensatz wird dann mittels Häufigkeitsverteilungen einzelner Merkmale oder Merkmalsgruppen oder auch spezifischer multivariater statistischer Methoden wie der Faktoranalyse oder durch Clustering analysiert. So können Texte und Register anhand der Dimensionen charakterisiert werden und Unterschiede aber auch Gemeinsamkeiten zwischen Registern festgestellt werden. Wissenschaftliche Texte werden z.B. als inhaltsbezogen (*informational*), nicht narrativ (*non-narrative*), abstrakt sowie als sorgfältig ausgearbeitet (*elaborate*) be-

¹Der Ansatz ist prinzipiell sprachunabhängig. Der englische Bezug ist der Tatsache geschuldet, dass die hier zitierte Registerforschung am Englischen durchgeführt wurde.

Summary of the co-occurrence patterns underlying five major dimensions of English.

DIMENSION 1 (Informational vs. Involved)		DIMENSION 2 (Narrative versus Non-Narrative)		DIMENSION 3 (Elaborated vs. Situating Reference)		DIMENSION 4 (Overt Expression of Persuasion)		DIMENSION 5 (Abstract versus Non-Abstract Style)	
nouns	0.80	past tense verbs	0.90	WH relative clauses on		infinitives	0.76	conjuncts	0.48
word length	0.58	third person pronouns	0.73	object positions	0.63	prediction modals	0.54	agentless passives	0.43
prepositional phrases	0.54	perfect aspect verbs	0.48	pied piping		suasive verbs	0.49	past participial	
type / token ratio	0.54	public verbs	0.43	constructions	0.61	conditional		clauses	0.42
attributive adjs.	0.47	synthetic negation	0.40	WH relative clauses on		subordination	0.47	BY-passives	0.41
		present participial		subject position	0.45	necessity modals	0.46	past participial	
private verbs	-0.96	clauses	0.39	phrasal coordination	0.36	split auxiliaries	0.44	WHIZ deletions	0.40
that deletions	-0.91			nominalizations	0.36	possibility modals	0.37	other adverbial	
contractions	-0.90	present tense verbs	-0.47					subordinators	0.39
present tense verbs	-0.86	attributive adjs.	-0.41	time adverbials	-0.60	[No complementary features]		[No complementary features]	
2nd person pronouns	-0.86			place adverbials	-0.49				
do as pro-verb	-0.82			other adverbs	-0.46				
analytic negation	-0.78								
demonstrative									
pronouns	-0.76								
general emphatics	-0.74								
first person pronouns	-0.74								
pronoun <i>it</i>	-0.71								
<i>be</i> as main verb	-0.71								
causative									
subordination	-0.66								
discourse particles	-0.66								
indefinite pronouns	-0.62								
general hedges	-0.58								
amplifiers	-0.56								
sentence relatives	-0.55								
WH questions	-0.52								
possibility modals	-0.50								
non-phrasal									
coordination	-0.48								
WH clauses	-0.47								
final prepositions	-0.43								

Abbildung 5.1: Die fünf Hauptdimensionen sprachlicher Varianz des Englischen nach Biber (1992: 335).

zeichnet, während Unterhaltungsliteratur als handlungsbezogen (*involved*), narrativ und nicht abstrakt angesehen wird.

Wie bereits erwähnt, ist die MD die bekannteste Methode zur Untersuchung von Registervariation. Allerdings müssen die sprachlichen Merkmale für die Untersuchung im Voraus definiert werden. Methoden aus dem Data Mining hingegen ermöglichen einen explorativen Ansatz zur Untersuchung von sprachlicher Varianz ohne vorher festgelegte Merkmale. Degaetano u.a. (2016) schlagen eine Methode vor, die von der Informationstheorie (Shannon 1949) inspiriert wurde. Sie basiert auf Sprachmodellen und verwendet die *Relative Entropie* (Kullback-Leibler Divergence, KLD) um sprachliche Merkmale zu finden, die im Sprachwandel, aber auch der Registervariation eine Rolle spielen.

Die Methoden erscheinen zunächst recht komplex, lassen sich aber durch interaktive Visualisierungen und webbasierte Korpora mit ebenfalls interaktiven Abfragewerkzeugen auch für die Fachdidaktik nutzbar machen. Im Folgenden wollen wir dies anhand eines Beispiels verdeutlichen, indem wir zunächst den informationstheoretischen Ansatz zur explorativen Analyse und Merkmalsfindung beschreiben und anschließend einen möglichen Einsatz von klassischen Korpuslinguistischen Methoden zur Merkmalsextraktion mit Frequenzanalyse vorstellen.

5.1.1 Registerforschung in der (universitären) Lehre

Das Unterrichten von korpuslinguistischen bzw. computerlinguistischen Methoden und der Verwendung der dazu benötigten Werkzeuge in geisteswissenschaftlichen Bereichen ist gleichermaßen eine Herausforderung wie eine Bereicherung. Auf der einen Seite stehen die Vorurteile und Bedenken der Teilnehmer und nicht zuletzt auch die Sinnfrage, auf der anderen Seite steht die zusätzliche Erkenntnisgewinn und die sich ergebenden Möglichkeiten und Erleichterungen. Dabei wird eines immer wieder deutlich: Sobald man die erste Hemmschwelle überwunden hat und auch der Nutzen sichtbar wird, werden die Methoden und Werkzeuge in der Regel sehr gut angenommen und führen auch recht schnell zu interessanten Ergebnissen.

Für die Einführung von Registervariation bieten wir in der universitären Lehre u.a. folgende Analysemethoden an:

- Merkmalsfindung und Merkmalsanalyse mit vorkompilierten explorativen interaktiven Visualisierungen von Registervergleichen auf Wortebene.
- Einfache Korpusabfragen mit anschließender Frequenzanalysen in online zugänglichen Korpusabfrage-Werkzeugen. Die Frequenzanalyse kann dabei auch, je nach Kenntnisstand der Teilnehmer, mit Programmen wie Excel oder R erfolgen.

Bei den Unterrichtseinheiten setzen wir so oft es geht aktives Lernen ein. Praktische Übungen, die im Unterricht allein oder in Gruppenarbeit bearbeitet werden und deren Ergebnisse dann in der Gruppe vorgestellt und gemeinsam diskutiert werden. Der Dozent wird zum Assistenten, der unterstützt, Hinweise gibt und versucht, zur selbstständigen Problemlösungsfindung zu verhelfen. Technische Probleme, die zum Teil auch bewusst vom Dozenten herbeigeführt wurden, können gemeinsam besprochen und gelöst werden. Wir haben festgestellt, dass die Studenten, je eigenständiger man sie arbeiten lässt, desto schneller potentielle Bedenken überwinden und desto besser den Lernstoff aufnehmen und verarbeiten. Interaktive Werkzeuge mit benutzerfreundlichen Oberflächen machen das Lernen interessant auch wenn man auf Abwege gerät.

5.1.2 Explorative Verfahren zur Merkmalsfindung und Merkmalsanalyse

Wie bereits erwähnt, werden in der Forschung Register zumeist anhand von vorher festgelegten Merkmalen untersucht. Methoden aus dem Data Mining ermöglichen aber auch einen eher explorativen Ansatz um Registervariation zu untersuchen. Dabei geht man nicht von vorher festgelegten Merkmalen aus, sondern versucht, die Merkmale mit Hilfe von Data Mining Methoden zu finden und zu analysieren. Interaktive Visualisierungen ermöglichen, auch ohne technisches oder sprachwissenschaftliches Vorwissen, einen nahezu barrierefreien Zugang zu diesen explorativen Methoden. Der visuelle Aspekt und die Interaktivität wecken das Interesse und ermöglichen einen unverkrampften Einstieg. Gleichzeitig bieten sie die Möglichkeit, aktives oder passives sprachliches Wissen anzuwenden und die Verbindung zwischen dem sprachlichen Signal (der Mikroebene) und übergeordneten sprachlichen Mustern und Prinzipien (der Makroebene) zu entdecken.

Die verwendeten Visualisierungen basieren auf vorkompilierten Data Mining Analysen. Sie zeigen ein bereits vom Text abstrahiertes Bild der sprachlichen Unterschiede, die Makroebene der sprachlichen Variation. Webbasierte Darstellungen mit interaktiven Schaltflächen erlauben verschiedene Blickwinkel auf die Analyse einzunehmen und zwischen verschiedenen Ebenen der Analyse zu wechseln, so auch auf die Mikroebene des konkreten sprachlichen Signals. Das Hauptaugenmerk kann auf die Analyse selbst gelegt werden.

Sehen wir uns dazu einen Vergleich verschiedener Register basierend auf der Wortebene an. Die Analyse basiert auf den Korpora der Brown Familie, die fünf verschiedene Korpora (BROWN, FROWN, BLOB, FLOB, LOB) und drei verschiedene Dimensionen sprachlicher Variation enthalten: Britisches vs. amerikanisches Englisch, Sprachwandel

A	press reportage	press
B	press editorial	press
C	press reviews	press
D	religion	general prose
E	skills.trades.hobbies	general prose
F	popular.lore	general prose
G	essay	general prose
H	misc	general prose
J	academic.learned	academic
K	general	fiction
L	adventure	fiction
M	science.fiction	fiction
N	mistery	fiction
P	romance	fiction
R	humor	fiction

Tabelle 5.1: Register in der Brown Corpus Familie.

repräsentiert durch drei Zeitperioden (1931, 1961, 1991) und Registervariation (Tabelle 1 zeigt einen Überblick der enthaltenen Register und Superregister).

Diese Korpora und ihre Subkorpora wurden auf der Basis von Unigrammen (Einwortsequenzen) verglichen und in einer webbasierten, interaktiven Visualisierung als Kombination aus Heatmaps und Wortwolken aufbereitet (s. Abbildung 2, Fankhauser u.a. 2014).

Die Register werden dabei mit Hilfe der relativen Entropie, der Kullbach-Leibler Divergence (KLD), verglichen. Sie vergleicht Sprachmodelle, als Verteilung von Wahrscheinlichkeiten (hier von Unigrammen, also Wörtern), in zwei verschiedenen Populationen (hier Registern). Die KLD berechnet die Anzahl zusätzlicher Bits, die für die Einkodierung von Einheiten einer Verteilung P basierend auf einer für die Verteilung Q optimierten Einkodierung benötigt werden als

$$D(Q||P) = \sum_x P(x) \log_2 \frac{P(x)}{Q(x)}$$

Dazu wird für jedes Register eine eigene Wahrscheinlichkeitsverteilung, ein Sprachmodell, erstellt. Für einen Vergleich werden dann die Register mit dem Sprachmodell des jeweilig anderen Registers enkodiert und das Resultat mit dem genuinen Sprachmodell verglichen. Gemessen wird die Anzahl zusätzlicher Bits, die für die Einkodierung benötigt werden. Je mehr zusätzliche Bits für die Einkodierung benötigt werden, desto

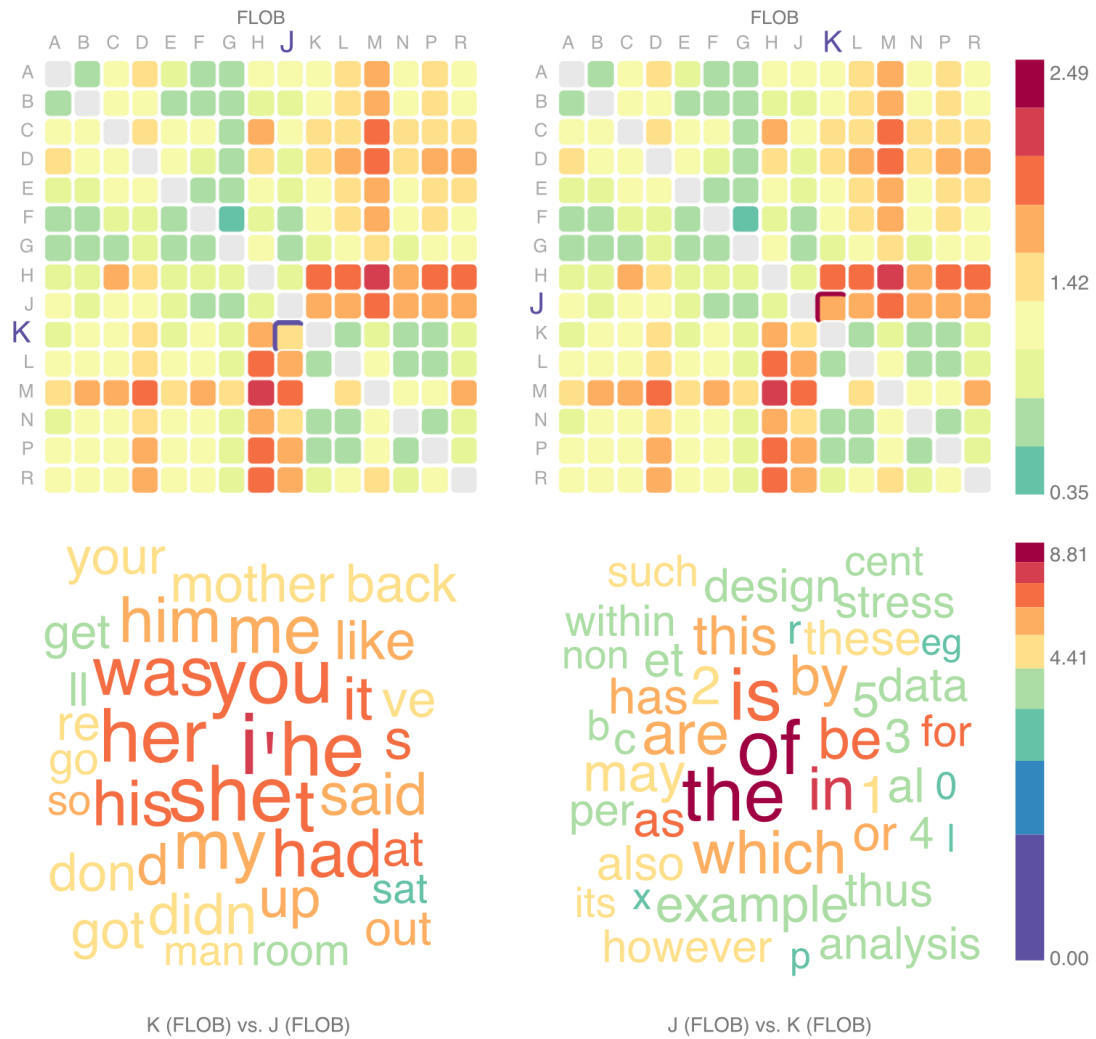


Abbildung 5.2: Visualisierung von Registervariation.

größer ist die Distanz bzw. desto größer der Unterschied zwischen den beiden Registern. Anschließend werden die gemessenen Distanzen dann in einer Heatmap visualisiert. Jedes Quadrat steht dabei für den Vergleich zwischen zwei Registern. Je (dunkel)roter die Farbe des Quadrats, desto mehr Bits werden für die Enkodierung gebraucht, desto größer ist die Distanz zwischen den beiden Registern, desto unterschiedlicher sind sie. Je grüner die Quadrate desto weniger zusätzliche Bits wurden benötigt, desto geringer der Unterschied.

Für eine genauere Untersuchung der Unterschiede können dann einzelne Quadrate ausgewählt werden. Für diese Auswahl wird eine Wortwolke gezeigt mit Merkmalen (hier Wörtern), die typisch für den jeweiligen Registervergleich sind. Sie basieren auf den KLD-Werten der einzelnen Wörter und einem Signifikanzmaß, dem *p-value*, und zeigen welche Wörter in einem Register typisch sind im Vergleich zu einem anderen Register. Auch hier gilt, je dunkelroter ein Wort ist, desto distinktiver, desto typischer ist es. Die Größe der Wörter zeigt die relative Häufigkeit des Wortes an, je größer das Wort, desto häufiger ist es.

Die Wortwolken in Abbildung 2 zeigen den Vergleich der Register J (wissenschaftliche Texte) und K (Unterhaltungsliteratur). Die linke Wortwolke zeigt typische Wörter für Unterhaltungsliteratur im Vergleich zu wissenschaftlichen Texten, die rechte Wortwolke typische Wörter für wissenschaftliche Texte im Vergleich zu Unterhaltungsliteratur.

Der technische bzw. mathematische Hintergrund klingt zunächst sehr komplex, er kann jedoch je nach Kenntnisstand unterschiedlich tief vermittelt werden. Für eine einfache Analyse der Darstellung reicht es folgende Punkte zu erklären:

- Vergleich basiert auf (Wahrscheinlichkeits)verteilung von Wörtern
- Heatmaps: Wahl des Registervergleichs über die Quadrate
- Farbschema (rot: distinktiv, grün: ähnlich)
- Wortwolke: zeigt typische Wörter
- Farbe: wie typisch, Größe: wie häufig

Für den ersten Einstieg belassen wir es auch in der universitären Lehre zumeist bei dieser minimalen Erklärung. Oft werden dann einzelne Details bei der gemeinsamen Analyse der Visualisierung erwähnt und erklärt. Eine systematische Aufbereitung der technischen und mathematischen Hintergründe erfolgt dann separat. Wir haben hier die Erfahrung gemacht, dass eine spätere Erklärung für das Verständnis der Methodik

eindrücklicher ist. Je nach Level der Studierenden werden Teile der Aufbereitung auch praktisch nachvollzogen und auf andere Merkmale angewandt. Für Lehrende ist zumindest ein grundlegendes Verständnis der Hintergründe wichtig, um auf potentielle Fragen und Rückfragen eingehen zu können. Je nach Interesse der SchülerInnen ist dann auch ein tiefer Einstieg möglich, eventuell auch fächerübergreifend angelegt, z.B. gemeinsam mit der Mathematik oder Informatik. Insbesondere die Mathematik, die in der Schule oft als zu abstrakt empfunden wird, könnte so den Bezug zu einem konkreten Einsatzfeld erhalten und ebenfalls profitieren.

Für die Analyse der Wortwolken in Abbildung 2 muss man zunächst von den Einzelwörtern abstrahieren, Gruppierungen vornehmen, Wörter oder Wortgruppen grammatisch klassifizieren (grammatische Kategorie, Wortklasse, etc.) und schließlich auf abstraktere linguistische Phänomene schließen.

Betrachtet man die Wortwolken unter diesen Gesichtspunkten, so ergibt sich z.B. Folgendes:

- Typisch für die Unterhaltungsliteratur sind Personalpronomen (*I, you, she, he*), Verben im past tense (*said, sat, got, was, had*).
- Typisch für wissenschaftliche Texte ist der bestimmte Artikel *the*, Präpositionen (vor allem *of*, aber auch *in, by, for, within*, Auxiliare im Präsens (*is, are, be, has*), Begriffe wie *example, analysis, design, data*.

Abstrahiert man weiter, so fällt auf, dass viele der typischen Wörter sogenannte Funktionswörter sind (Pronomen, Artikel, Auxiliare) und nur wenige Inhaltswörter (Verben, Nomen). Im klassischen Data Mining werden Funktionswörter oft als sogenannte *stop words* weggeworfen. Sie können durch ihren eher funktionalen Charakter jedoch wichtige Hinweise auf sprachliche Eigenschaften eines Textes geben.

Vergleichen wir unsere typischen Wörter nun mit den fünf Dimensionen des Englischen von Biber (1992), so ergeben sich für wissenschaftliche Texte folgende Eigenschaften:

- Fakten bezogen (*informational*), gekennzeichnet durch einen nominalen Stil (Nomen und Artikel bzw. Nominalphrasen) sowie die Verwendung von Präpositionalphrasen (Präpositionen)
- *non-narrativ*, gekennzeichnet durch die Verwendung von Präsensverben und Relativsätzen (*which*)

- *abstrakt*, durch die Verwendung von kompakteren Strukturen in Form von einfachen Satzstrukturen (X ist Y) und einem nominalen Stil

Halliday (1988) und Halliday u.a. (2005) sprechen hier von einer größeren Enkodierungsdichte in wissenschaftlichen Texten, die zur Effizienz der Kommunikation beiträgt. Der inhaltliche Aspekt wird zum großen Teil durch (komplexe) Nominalphrasen ausgedrückt, die Satzstruktur wird einfach gehalten mit eher inhaltsleeren Verben, wie *to be*, das hier als Vollverb fungiert. Bei der Unterhaltungsliteratur sehen wir Hinweise auf folgende Eigenschaften:

- *involved* (beteiligt), durch die Verwendung von Personalpronomen
- *narrativ*, durch die Verwendung von Verben im *past tense*, sowie Verben des Sagens (*said*) und Aktionsverben wie *setzen* oder *gehen*.
- direkte und indirekte Rede, gekennzeichnet durch das Wort *said* (sagte)

Einige der Rückschlüsse auf Eigenschaften und übergeordnete linguistische Phänomene sind zunächst nur Vermutungen, die belegt werden müssen. Dafür muss von der Makrostruktur auf die Mikrostruktur, das konkrete Sprachsignal, gewechselt werden. Die Visualisierung bietet hierfür ein interaktives Feature, das es erlaubt, durch das Klicken auf ein Wort in der Wortwolke zu den Konkordanzen (Textstellen) für das jeweilige Wort im Korpus zu springen. Die Visualisierung setzt so Verbindungen zu einer Repräsentation des Korpus auf einer online Abfrageplattform (CQPweb, Hardie 2011). Abbildung 3 zeigt die Konkordanz für *said* im Register Unterhaltungsliteratur.

Wir sehen auf den ersten Blick, dass es sich hier tatsächlich zumeist um Beispiele für direkte und indirekte Rede handelt, unsere diesbezügliche Annahme wird also bestätigt. Wir könnten jetzt zur Visualisierung zurückkehren und die nächste Annahme auf diese Weise überprüfen. Eine andere Möglichkeit ist, sich die Konkordanzen als konkrete Realisierungen für dieses Phänomen (näher) anzuschauen. Das Bewusstsein für den Gebrauch dieses Phänomens und für die sprachliche Variation wird so geschärft. Da die Anzeige der Konkordanzen im online Korpus erfolgt, können wir die Beispiele sortieren und gruppieren und typische Kontexte und Kookkurenzen² identifizieren.

Die explorative Methode der Merkmalsfindung und Merkmalsanalyse bietet für die Fachdidaktik viele Möglichkeiten. Es können zum einen registerspezifische sprachliche Merkmale herausgearbeitet werden, zum anderen vermittelt es gleichzeitig die Fähigkeit passives und/oder aktives sprachliches Wissen anzuwenden sowie vom sprachlichen

²Unter Kookkurenzen versteht man das gemeinsame Auftreten.

No	Filename	Solution 1 to 50	Page 1 / 5
1	K01_1	' was , it seemed , a way husbands had . She said , ' I am married , too . I live in this	
2	K01_1	Then Dion had a quiet talk with his wife . He said , ' Look here , Polly dear , this can ' t	
3	K02_1	week . Maideley and I are old friends , you know ' said the Bishop , as she went to take her place on the	
4	K02_1	in his study . A very charming girl your sister , ' said the Bishop to Jim , very charming . ' Others near who	
5	K02_1	near by . I ' ll bring tea here , ' he said . It ' s cool and quiet . ' In a moment	
6	K02_1	air of being amused with himself . And you ? ' he said . Eastbourne was only an accident with you , wasn ' t	
7	K02_1	London cannot be compared with a place like this , ' he said after a moment ' s pause , deliberately , as if in	
8	K02_1	most people , lose their attraction when they are known , he said quietly . It ' s best to keep the guilt on our	
9	K02_1	I don ' t think life is like that , ' she said at length . I believe it is beautiful at the core .	
10	K02_1	You are happy to be able to think so , ' he said , watching her closely , still with his disconcerting smile . But	
11	K02_1	ways . Perhaps you can supply a better one , ' he said . I ' m not wedded to it . I just threw	
12	K02_1	her quick impulse to anger had gone . Mine ? ' she said in response . I don ' t think I have one ;	
13	K02_1	must get used to people who did not mean exactly what they said . There was the walk home afterwards over the hills and fields	
14	K03_1	Why not ? " . " Well , when the best is said about it , " I replied , " it is only a	
15	K03_1	never called him anything , " I replied . " I simply said I met a Methodist minister . " She blushed violently . "	
16	K03_1	path . " Don ' t I feel virtuous ! " she said , as we neared the old church porch . Why ? "	
17	K03_1	too , but for father . As one of the chapel people said to me the other day , our rector believes in salvation by	
18	K03_1	village street by Mr Bunyan ' s side . As I have said , although more than twenty one years of age , I had	
19	K03_1	doubt as to what they ought to do . As I have said , there was scarcely any light , especially in the back part	
20	K03_1	he appealed to me as a striking figure . As I have said earlier in these pages , when I had met him first he	
21	K03_1	Where I sat , indeed , was , as I have said , in deep gloom , as was a great part of the	
22	K04_1	for instance . Now that/IN he was himself a man , said Edmund , it was time that/IN something should be done about this	
23	K04_1	re back in London management again . " Yes , " " said she , and her dark eyes flashed from their dark hollows as	
24	K04_1	be rude . Of course your experience is invaluable . You always said when I was twenty one you would hand over the whole concern	

Abbildung 5.3: Konkordanzen für 'said' im Register Unterhaltungsliteratur.

Signal zu abstrahieren und auf übergeordnete Phänomene zu schließen. So müssen etwa grammatische Kategorien (Wortarten) erkannt und auf übergeordnete Eigenschaften von Texten bezogen werden. Je nach Wissensstand kann dies mit mehr oder weniger Unterstützung durch das Lehrpersonal erfolgen. Oft genügen hier kleine Beispiele, um das Prinzip zu veranschaulichen. So lassen sich in einer Wortwolke immer wieder neue Phänomene erkennen, die auch der Lehrende noch nicht entdeckt hat. Durch die Möglichkeit, sich die Konkordanzen für ein bestimmtes Wort anzeigen zu lassen, bleibt der Bezug zum konkreten sprachlichen Signal erhalten und man erhält Beispiele für den tatsächlichen Gebrauch.

5.2 Merkmalsextraktion und Häufigkeitsverteilung

Die klassische Methode für die Untersuchung von Registervariation ist die Merkmalsextraktion mit anschließender Häufigkeitsverteilung, wie sie auch für die multidimensionale Analyse von Biber (1992) angewendet wird. Wir haben uns trotzdem dazu entschlossen, sie nicht zuerst zu betrachten, weil das explorative Verfahren einen intuitiveren und zunächst einfacheren Zugang ermöglicht. Für die klassische Methode benötigt man zumindest Grundkenntnisse in korpuslinguistischen Methoden: Syntax einer Abfragesprache bzw. einfache reguläre Ausdrücke, Fähigkeit ein linguistisches Phänomen zu formalisieren, relative Häufigkeit bzw. Häufigkeitsverteilung. Die Vorbereitung mit den Studierenden ist hier also etwas zeitaufwendiger.

Ein weiterer Grund, mit dem explorativen Verfahren anzufangen, ist, dass die Merkmalsextraktion auch ein logischer nächster Schritt in der Analyse darstellt. Zuerst werden die Merkmale mit einem explorativen Verfahren identifiziert und dann mit korpuslinguistischen Mitteln zunächst formalisiert, dann extrahiert und analysiert.

Nehmen wir als Beispiel eines der sprachlichen Merkmale, die wir mit der explorativen Methode herausgearbeitet haben, die Alternation der Verbtempi *past tense* (Unterhaltungsliteratur) und *present tense* (wissenschaftliche Texte). Allerdings extrahieren wir nicht selektiv nur diese beiden Verbtempi für dieses Registerpaar, sondern alle Verformen im Korpus.

Für die Extraktion verwenden wir wieder die Online-Version der Brown Corpora in CQPweb (Hardie 2012). Die Plattform erlaubt einen relativ einfachen und intuitiven Zugang zur Korpusextraktion. Sie basiert auf der Abfragesprache CQP (Evert u.a. 2011) und macht somit auch komplexere Abfragen möglich. Einige nützliche statistische Auswertungen sind in der Plattform selbst möglich (Frequenzanalyse, Häufigkeitsverteilung, Kollokationsextraktion, etc.). Die Ergebnisse können aber auch als multivariater Daten-

The Brown corpus family: *powered by CQPweb*

Standard Query

[pos="VV.*"]

Query mode: CQP syntax ▼ [Sample Queries in CQP syntax](#)
[Simple query language syntax](#)

Number of hits per page: 50 ▼

Restriction: None (search whole corpus) ▼

Start Query
Reset Query

Abbildung 5.4: CQPweb Eingabefenster mit Abfrage für Vollverben in CQP Syntax.

satz im TAB-getrennten Format heruntergeladen werden. Das Format ist ein Standard Austauschformat und kann in vielen Programmen für die weitere statistische Auswertung benutzt werden, z.B. in Excel oder R.

Die Syntax der Abfragesprache basiert auf regulären Ausdrücken und bedarf zumindest einer kurzen Einführung. Es gibt zwei Syntaxmodi, die CQP Syntax und eine „vereinfachte“ Syntax. Für die Extraktion aller Vollverben aus einem Korpus lautet die Abfrage:

- in CQP Syntax: [pos=„VV.*“]
- in der „einfachen“ Syntax: _VV* #

Abbildung 4 zeigt das CQPweb Eingabefenster mit der Abfrage für Vollverben in der CQP Syntax.

Das Resultat dieser Abfrage ist zunächst unsortiert und zeigt alle Vorkommen von Vollverben im Korpus (s. Abbildung 5).

Konkordanzen sind, wie auch andere Ansichten in CQPweb, interaktiv. Mit dem Mauszeiger kann man sich z.B. für einzelne Ergebnisse die Wortarten anzeigen lassen,

Your query "[pos="VV.*"]" returned 580,106 matches in 2,471 different texts (in 5,895,588 words [2,471 texts]; frequency: 98,396.63 instances per million words) [0.113 seconds - retrieved from cache]

Navigation: |< << >> >| Show Page: 1 Line View Show in random order New query Go!

No	Filename	Solution 1 to 50	Page 1 / 11603
1	bl_A01_1	LONDON . NEW MEETING OF CONFERENCE IN THE AUTUMN . Mr Baldwin	Supports Irwin And Defies Churchill . HISTORIC DEBATE . THE final decisive meetings
2	bl_A01_1	MEETING OF CONFERENCE IN THE AUTUMN . Mr Baldwin Supports Irwin And	Defies Churchill . HISTORIC DEBATE . THE final decisive meetings of the India
3	bl_A01_1	THE final decisive meetings of the India Round Table Conference will be	held in London in the early autumn . This announcement was made by
4	bl_A01_1	be held in London in the early autumn . This announcement was	made by Mr Wedgwood Benn , the Secretary for India , during the
5	bl_A01_1	course of an historic debate in the House of Commons . In	closing the debate , the Prime Minister stated that/IN Mr Gandhi and other
6	bl_A01_1	House of Commons . In closing the debate , the Prime Minister	stated that/IN Mr Gandhi and other Congress leaders will almost certainly come to
7	bl_A01_1	Minister stated that/IN Mr Gandhi and other Congress leaders will almost certainly	come to this country for the resumed Round Table Conference . The debate
8	bl_A01_1	other Congress leaders will almost certainly come to this country for the	resumed Round Table Conference . The debate was made tense by a fierce
9	bl_A01_1	this country for the resumed Round Table Conference . The debate was	made tense by a fierce duel between Mr Baldwin and Mr Winston Churchill
10	bl_A01_1	by a fierce duel between Mr Baldwin and Mr Winston Churchill which	threatens to split the Tory Party in twain . Mr Baldwin made a
11	bl_A01_1	fierce duel between Mr Baldwin and Mr Winston Churchill which threatens to	split the Tory Party in twain . Mr Baldwin made a passionate speech
12	bl_A01_1	which threatens to split the Tory Party in twain . Mr Baldwin	made a passionate speech defending Party co operation to secure an India settlement
13	bl_A01_1	the Tory Party in twain . Mr Baldwin made a passionate speech	defending Party co operation to secure an India settlement and challenging the diehards
14	bl_A01_1	Mr Baldwin made a passionate speech defending Party co operation to	secure an India settlement and challenging the diehards in his own Party .
15	bl_A01_1	passionate speech defending Party co operation to secure an India settlement and	challenging the diehards in his own Party . If " they are in
16	bl_A01_1	Party . If " they are in a majority , " he	said , in " God ' s name let them choose a man
17	bl_A01_1	majority , " he said , in " God ' s name	let them choose a man to lead them . " The Churchill faction
18	bl_A01_1	" he said , in " God ' s name let them	choose a man to lead them . " The Churchill faction must now
19	bl_A01_1	in " God ' s name let them choose a man to	lead them . " The Churchill faction must now destroy Mr Baldwin '
20	bl_A01_1	a man to lead them . " The Churchill faction must now	destroy Mr Baldwin ' s position , or risk expulsion from the Tory

Abbildung 5.5: Konkordanzen Vollverben.

durch Klicken auf die Ergebnisverben wird mehr Kontext gezeigt und über den Dateinamen erhält man zusätzliche Information über den Text, in dem das Verb vorkommt.

Weitere Auswertungen sind über ein Menü zugänglich. Interessant ist hier die Häufigkeitsverteilung über die verschiedenen Subkorpora (etwa die Register) oder die Häufigkeitszählung der Verbformen bzw. der dazugehörigen Wortarten. Abbildung 6 zeigt die Häufigkeiten der verschiedenen Verbwortarten im FLOB Corpus.

Your query "[pos="VV.*"]" returned 116,243 matches in 1,585 different texts (in 1,163,712 words [1,587 texts]; frequency: 99,889.84 instances per million words)

Showing frequency breakdown of annotation in this query, at the query node; there are 6 different types and 116,243 tokens at this concordance position.

Navigation: |< << >> >| Breakdown position: Node Frequency breakdown of annotation only Go!

No.	Search result	No. of occurrences	Percent
1	VVN	28234	24.29%
2	VV	27413	23.58%
3	VVD	25724	22.13%
4	VVG	17840	15.35%
5	VVZ	8619	7.41%
6	VVP	8413	7.24%

Abbildung 5.6: Häufigkeiten der Wortarten der Vollverben in FLOB.

Die Wortarten werden im Korpus durch Wortartentags repräsentiert, die durch ein Tagset (hier das UPenn Tagset, Marcus u.a. 1993) definiert sind. Die folgende Tabelle 2 zeigt, für welche Wortarten die einzelnen Tags in Abbildung 6 stehen.

VV	verb, base form
VVD	verb, past tense
VVG	Verb, gerund or pres. part.
VVN	verb, past participle
VVP	Verb, non-3rd p sing. pres.
VVZ	Verb, 3rd person sing. pres.

Tabelle 5.2: Wortartentags von Vollverben.

Die häufigste Wortart im FLOB Korpus ist demnach das *past participle*, gefolgt von Verben in der Grundform und im *past tense*. Wir sehen, dass auch diese Ansicht interaktiv ist. So können wir uns ganz einfach die Konkordanzen zu einer der Wortarten anzeigen lassen, z.B. zum *past tense*. Eine Häufigkeitsverteilung der *past tense* Verbformen zeigt folgende Distribution über die Register (Abbildung 7).

Das Fenster für die Häufigkeitsverteilung zeigt für jedes Register die Gesamtanzahl der Wörter, die Anzahl der Treffer, die Verteilung der Treffer über die verschiedenen Texte, sowie die relative Häufigkeit der Treffer. Vergleichen wir hier die relativen Häufigkeiten, so wird schnell deutlich, dass das *past tense* vor allem in den verschiedenen Literaturgenres häufig vorkommt, wenn auch weniger ausgeprägt im Register „Humor“. In wissenschaftlichen Texten kommt der *past tense* im Vergleich hingegen selten vor. Die Vermutung, die wir aufgrund der explorativen Analyse getroffen haben, wird durch die Häufigkeitsverteilung also bestätigt.

Auch diese Ansicht ist interaktiv. So kann man sich zu jedem Register wieder die entsprechenden Konkordanzen anzeigen und anschließend z.B. Häufigkeiten für die Wortformen ausgeben lassen. Die folgende Tabelle zeigt die 15 häufigsten Verbformen für die Register wissenschaftliche Texte und Unterhaltungsliteratur.

Wir sehen, dass es hier kaum Überschneidungen gibt. Sowohl für die Unterhaltungsliteratur als auch für wissenschaftliche Texte lassen sich semantische Gruppen identifizieren. Bei wissenschaftlichen Texten sind es vor allem Fakten und inhaltsbezogene Verben wie *showed*, *described*, *found*, *provided*, *used*, *reported*, während es bei der Unterhaltungsliteratur narrative, aktionsbezogene Verben oder Verben des Sagens sind wie *said*, *looked*, *thought*, *went*, *came*.

Die Interaktionsmöglichkeiten in CQPweb sind vielfältig und man kann fast beliebig zwischen Gruppierungen, Häufigkeitsverteilungen und Konkordanzen wechseln. So

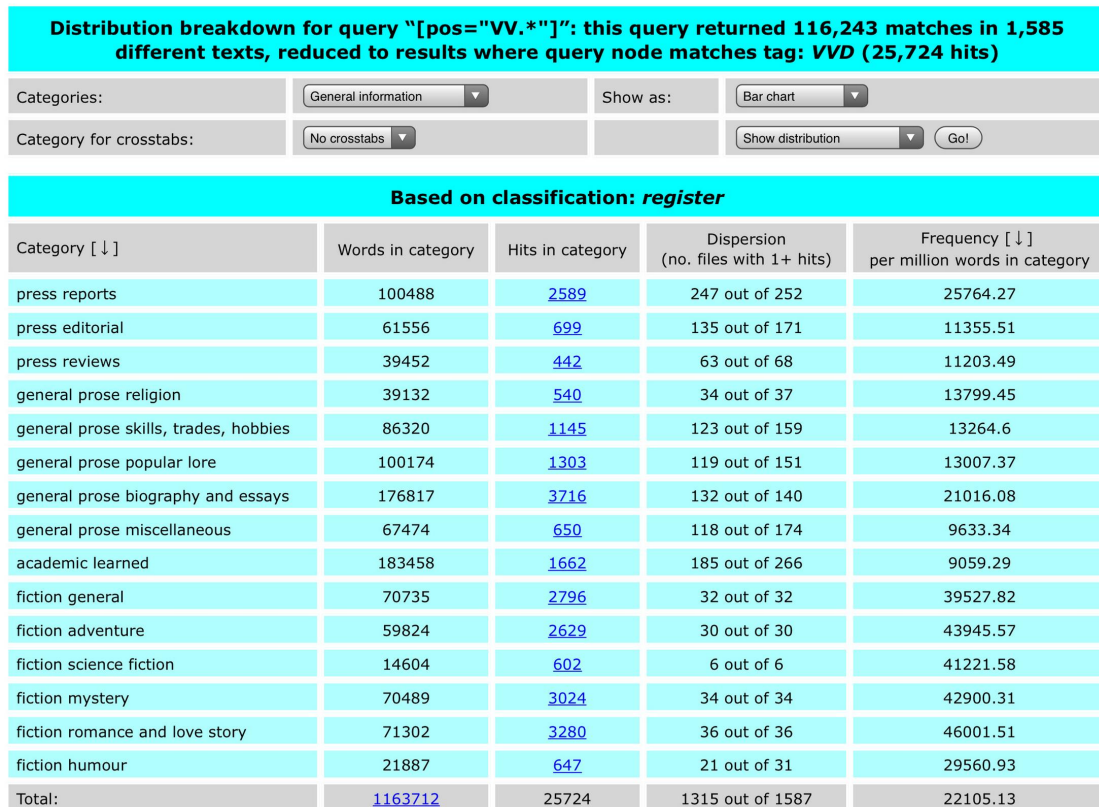


Abbildung 5.7: Häufigkeitsverteilung des past tense über die Register.

wird zum einen das Zusammenspiel zwischen Makro- und Mikroanalyse deutlich und zum anderen verliert man nie den Blick auf das tatsächliche sprachliche Signal. Auch Kollokationen können als Generalisierung über den Kontext herangezogen werden. So erlaubt auch CQPweb einen explorativen Umgang mit den Ergebnissen einer Abfrage. Dabei wird auch hier die Fähigkeit vermittelt, sprachliches Wissen anzuwenden und vom sprachlichen Signal und den Auswertungen zu abstrahieren bzw. sie miteinander in Verbindung zu setzen.

Wir erinnern uns an das Beispiel aus dem vorherigen Kapitel, mit den Konkordanzen zu der Verbform *said* in der Unterhaltungsliteratur. Es handelt sich zumeist um Beispiele für direkte und indirekte Rede. Schauen wir uns die Instanzen von *said* in wissenschaftlichen Texten an, so finden wir nur vereinzelt direkte Rede (zumeist in populärwissenschaftlichen Texten oder bei Zitaten). Wir finden indirekte Rede bei paraphrasierten Referenzen und recht häufig eine sehr spezifische Konstruktion *be said to*, die wir in der Unterhaltungsliteratur nicht finden. Auch bei anderen Verbformen lassen

Wissenschaftliche Texte			Unterhaltungsliteratur		
Verbform	Anzahl	Prozent	Verbform	Anzahl	Prozent
did	99	5.96	did	212	7.58
showed	35	2.11	said	187	6.69
became	31	1.87	looked	71	2.54
made	26	1.56	thought	61	2.18
described	25	1.5	went	60	2.15
found	24	1.44	came	55	1.97
took	24	1.44	knew	52	1.86
gave	19	1.14	seemed	50	1.79
said	19	1.14	got	50	1.79
provided	18	1.08	made	46	1.65
used	17	1.02	told	45	1.61
reported	17	1.02	took	40	1.43
produced	16	0.96	found	37	1.32
came	15	0.9	saw	35	1.25
went	15	0.9	asked	33	1.18

Tabelle 5.3: Häufigste Wortformen für past tense Verben.

sich für wissenschaftliche Texte Muster finden, die häufiger auftreten, als dies bei der Unterhaltungsliteratur der Fall ist, so z.B. *it came to, became + Adjektiv*.

CQPweb bietet durch seine interaktiven Schaltflächen gute Möglichkeiten, sprachliche Variationen genauer zu untersuchen und dabei zwischen verschiedenen Abstraktionsebenen (Makro- und Mikroebene) zu wechseln. Durch die Möglichkeit Ergebnisse und deren Kontext zu sortieren, zu gruppieren und zu analysieren, können hier die feineren Mechanismen sprachlicher Variation (explorativ) untersucht werden.

Für einen Vergleich aller Wortarten in einem Balkendiagramm muss man allerdings die Ergebnisse exportieren und in einem externen Statistikprogramm bearbeiten. CQPweb kann Daten in einem TAB-getrennten Format (s. Tabelle 5) exportieren, so dass für die weitere Auswertung viele Programme in Frage kommen, u.a. auch Excel oder R.

Für einen ersten Einstieg scheint Excel zunächst einfacher, erfordert aber einige manuelle Schritte, bis man ein Balkendiagramm mit relativen Häufigkeiten erstellt hat. R erscheint zunächst weniger geeignet für Einsteiger. Hier kann man sich aber mit einem vordefinierten R-Markdown Dokument behelfen, das es auch Anfängern erlaubt die Auswertung in R durchzuführen. Ein solches Dokument definiert alle notwendigen R-Befehle vor. Es muss nur noch ausgeführt werden, wofür es inzwischen mit RStudio benutzerfreundliche Hilfsprogramme gibt. Das Ergebnis des R-Markdown Dokuments kann man sich in verschiedenen Formaten (HTML, Word, PDF) ausgeben lassen. Es enthält ne-

repeated	repeat	VVD	A
said	say	VVD	A
said	say	VVD	A
said	say	VVD	A
demonstrated	demonstrate	VVD	A
argued	argue	VVD	A
said	say	VVD	A
suffered	suffer	VVD	A
won	win	VVD	A
dipped	dip	VVD	A
held	hold	VVD	A
won	win	VVD	A
took	take	VVD	A
took	take	VVD	A
expected	expect	VVD	A

Tabelle 5.4: TAB-getrenntes Exportformat.

ben Informationen zur Funktion auch die fertigen Schaubilder. Eine gute Möglichkeit ist hier, in Gruppen arbeiten zu lassen, so dass Probleme gemeinsam angegangen und gelöst werden können.

Abbildung 8 zeigt die Häufigkeitsverteilung der Wortartentags von Vollverben in der Brown Corpus Familie. Die Abbildung zeigt für alle Literaturgattungen (K-R) eine relativ ähnliche Verteilung der verschiedenen Verbformen. Das *past tense* ist jeweils die häufigste Verbform. Bei wissenschaftlichen Texten hingegen ist das *past tense* eher selten, es überwiegt das *present perfect*. Von den anderen Registern ist nur noch bei *press reportage* die Verbform *past tense* am häufigsten. Es fällt auch auf, dass von den Supergenres (*press, prose, academic, fiction*) die literarischen Register am homogensten sind.

Wie bei den anderen Analysen muss auch hier sprachliches Wissen angewandt und müssen abstrakte Kategorien interpretiert werden. Die Auswertungen mit externen statistischen Programmen benötigen trotz aller Vorbereitung mehr Einarbeitungszeit. In einem schulischen Kontext sind sie wohl am ehesten in Verbindung mit einem fachübergreifenden Projekt mit der Mathematik oder „Informatik“(-AG) realisierbar.

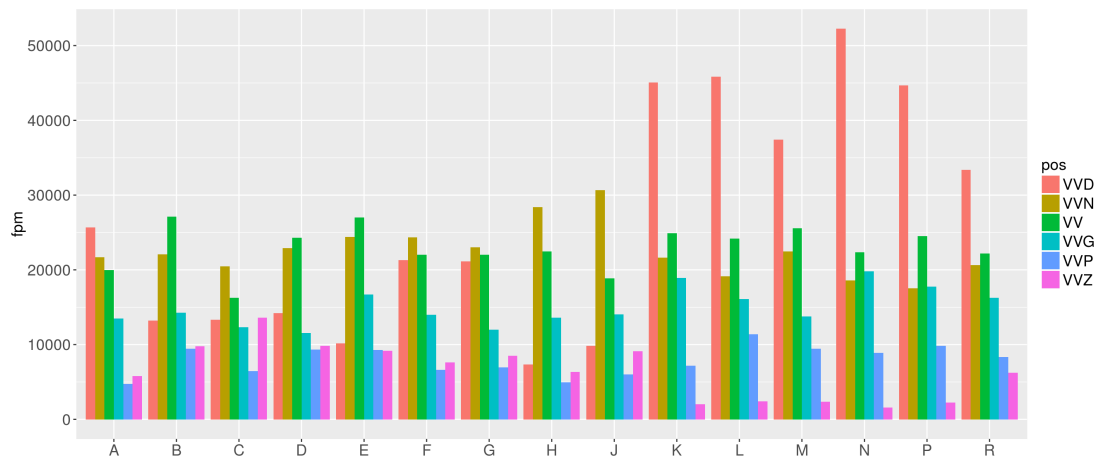


Abbildung 5.8: Häufigkeitsverteilung der Wortartentags von Vollverben.

5.3 Zusammenfassung

Wir haben zwei verschiedene, aber kombinierbare Methoden aus der Korpuslinguistik und Registerforschung vorgestellt. Dabei ging es vor allem um die Anwendbarkeit und den potentiellen Nutzen in der Fachdidaktik.

Wir haben gezeigt, dass interaktive, webbasierte Visualisierungen einen intuitiven und explorativen Zugang zu sprachlicher Variation bieten können. Auch ohne technische oder mathematische Vorkenntnisse kann mit diesen Methoden textbasiert sprachliche Variation erfahrbar gemacht werden. Dabei wird nicht nur das Bewusstsein für Registervariation und die damit verbundenen sprachlichen Mittel geschärft, sondern auch das Anwenden von aktivem oder passivem sprachlichen Wissen und zudem die Fähigkeit zu abstrahieren vermittelt.

Wir haben auch gezeigt, dass textbasierte Methoden durch ihre Nähe zum realen, nicht konstruierten, sprachlichen Signal die Verwendung von typischen und spezifischen sprachlichen Mustern verdeutlichen können. Gerade der Wechsel zwischen Makro- und Mikroebene ist wichtig, um Registerbewusstsein sowohl in der Perzeption als auch in der Produktion zu schärfen. Gleichzeitig wird die Fähigkeit trainiert aktives aber vor allem auch passives sprachliches Wissen anzuwenden.

Literaturhinweise

- Biber, Douglas (1988): *Variation Across Speech and Writing*, Cambridge.
- Biber, Douglas (1992): The Multi-Dimensional Approach to Linguistic Analyses of Genre Variation. An Overview of Methodology and Findings, in: *Computers and the Humanities* 26.5-6, 331-345.
- Biber, Douglas (1995): *Dimensions of Register Variation. A Cross-Linguistic Comparison*, Cambridge.
- Biber, Douglas u.a. (2001): Register variation. A corpus approach, *The handbook of discourse analysis*, 175-196.
- Biber, Douglas u.a. (1999): *Longman Grammar of Spoken and Written English*, Harlow.
- Degaetano-Ortlieb, Stefania u.a. (2016): An Information-Theoretic Approach to Modeling Diachronic Change in Scientific English, in: Suhr, Carla u.a. (Hg.): *Selected Papers from Varieng. From Data to Evidence (d2e)*, Helsinki.
- Evert, Stefan u.a. (2011): Twenty-First Century Corpus Workbench. Updating a Query Architecture for the New Millennium, in *Proceedings of the Corpus Linguistics 2011 Conference*, Birmingham.
- Fankhauser, Peter u.a. (2014): Exploring and Visualizing Variation in Language Resources, in: Calzolari, Nicoletta u.a. (Hg.): *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik.
- Halliday, Michael A. K. (1985): *An introduction to functional grammar*, New York.
- Halliday, Michael A. K. (1988): On the Language of Physical Science, in: Ghadessy, Mohsen (Hg.): *Registers of Written English. Situational Factors and Linguistic Features*, London, 162-177.
- Halliday, Michael A. K. u.a. (1985): *Language, Context, and Text. Aspects of Language in a Social-semiotic Perspective*, Oxford.
- Halliday, Michael A. K. u.a. (2005): *Writing Science. Literacy and Discursive Power*, London u.a.
- Hardie, Andrew (2012): CQPweb – Combining Power, Flexibility and Usability in a Corpus Analysis Tool, in: *International Journal of Corpus Linguistics* 17.3, 380-409.

- Marcus, Mitchell P. u.a. (1993): Building a large annotated corpus of English. The Penn Treebank, in: Computational Linguistics 19.2, 313-330.
- Quirk, Randolph u.a. (1985): A comprehensive grammar of the English language, London.
- Shannon, Claude E. (1949): The mathematical theory of communication. Urbana u.a.
- Zwicky, Arnold M. u.a. (1982): Register as a dimension of linguistic variation. Sublanguage, Studies of Language, in: Restricted Semantic Domains 9, 213-218.